The Annals of Mathematical Statistics
1970, Vol. 41, No. 6, 1906–1916

SOME REMARKS ON THE TWO-ARMED BANDIT[1]

BY J. FABIUS AND W. R. VAN ZWET

University of Leiden and Mathematisch Centrum

1. Introduction and summary. In this paper we consider the following situation: An experimenter has to perform a total of N trials on two Bernoulli-type experiments E_1 and E_2 with success probabilities α and β respectively, where both α and β are unknown to him. The trials are to be carried out sequentially and independently, except that for each trial the experimenter may choose between E_1 and E_2, using the information obtained in all previous trials. The decisions on the part of the experimenter to use E_1 or E_2 in the successive trials may be randomized, i.e. for any trial he may use a chance mechanism in order to choose E_1 or E_2 with probabilities δ and $1-\delta$ respectively, where δ may depend on the decisions taken and the results obtained in the previous trials. A strategy Δ will be a set of such δ's, completely describing the experimenters behavior in every conceivable situation.

We assume the experimenter wants to maximize the number of successes. More precisely, we assume that he incurs a loss

$$L(\alpha, \beta, s) = N \max(\alpha, \beta) - s \tag{1.1}$$

if he scores a total of s successes. If he uses a strategy Δ, his expected loss is then given by the risk function

$$R(\alpha, \beta, \Delta) = N \max(\alpha, \beta) - E(S \mid \alpha, \beta, \Delta), \tag{1.2}$$

where S denotes the random number of successes obtained. Thus the risk of a strategy Δ equals the expected amount by which the number of successes the experimenter will obtain using Δ falls short of the number of successes he would score if he were clairvoyant and would use the more favorable experiment throughout the N trials. It is easy to see that $R(\alpha, \beta, \Delta)$ also equals $|\alpha-\beta|$ times the expected number of trials in which the less favorable experiment is performed under Δ.

We say that state $(m, k; n, l)$ is reached during the series of trials if in the first $m+n$ trials E_1 is performed m times, yielding k successes, and E_2 is performed n times, yielding l successes. Clearly, under a strategy Δ, the probability that this will happen is of the form

$$\pi_{\alpha,\beta,\Delta}(m, k; n, l) = p_\Delta(m, k; n, l)\alpha^k(1-\alpha)^{m-k}\beta^l(1-\beta)^{n-l}, \tag{1.3}$$

where $p_\Delta(m, k; n, l)$ depends on the state $(m, k; n, l)$ and the strategy Δ, but not on α and β. It is easy to show (e.g. by induction on N) that the class of all strategies is convex in the sense that there exists, for every pair of strategies Δ_1 and Δ_2 and for every $\lambda \in [0, 1]$, a strategy Δ such that

$$p_\Delta(m, k; n, l) = \lambda p_{\Delta_1}(m, k; n, l) + (1-\lambda) p_{\Delta_2}(m, k; n, l) \tag{1.4}$$

Received November 1, 1968; revised September 3, 1969.

[1] Report S-399, Mathematisch Centrum, Amsterdam.

S. van de Geer and M. Wegkamp (eds.), *Selected Works of Willem van Zwet*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1314-1_2, © Springer Science+Business Media, LLC 2012

for every state $(m, k; n, l)$. Moreover, this strategy Δ can always be taken to be such, that according to it the experimenter should base all his decisions exclusively on the numbers of successes and failures observed with E_1 and E_2, irrespective of the order in which these data became available. Denoting the class of all such strategies by $\mathscr{D}$ and remarking that $R(\alpha, \beta, \Delta)$ can be expressed in terms of the $\pi_{\alpha,\beta,\Delta}(m, k; n, l)$, we may conclude that $\mathscr{D}$ is an essentially complete class of strategies. We denote the probabilities δ constituting any strategy in $\mathscr{D}$ by $\delta(m, k; n, l)$: the probability with which the experimenter, having completed the first $m+n$ trials and thereby having reached state $(m, k; n, l)$, chooses E_1 for the next trial.

We note that if $p_\Delta(m, k; n, l) = 0$ for a state $(m, k; n, l)$, then $\delta(m, k; n, l)$ does not play any role in the description of Δ and may be assigned an arbitrary value without affecting the strategy. We shall say that any strategy Δ' such that $p_{\Delta'}(m, k; n, l) = p_\Delta(m, k; n, l)$ for all states $(m, k; n, l)$ constitutes a version of Δ.

Since we are considering a symmetric problem in the sense that it remains invariant when α and β are interchanged, it seems reasonable to consider strategies with a similar symmetry. Thus we are led to define the class $\mathscr{L}$ of all symmetric strategies: $\Delta \in \mathscr{L}$ iff $\Delta \in \mathscr{D}$ and $\delta(m, k; n, l) = 1-\delta(n, l; m, k)$ for all states $(m, k; n, l)$ with $p_\Delta(m, k; n, l) \neq 0$. Clearly, for $\Delta \in \mathscr{L}$,

$$\delta(m,k;m,k) = \tfrac{1}{2} \quad \text{if} \quad p_\Delta(m,k;m,k) \geqq 0, \tag{1.5}$$

and

$$p_\Delta(m,k;n,l) = p_\Delta(n,l;m,k) \quad \text{for all states} \quad (m,k;n,l). \tag{1.6}$$

It follows that, for $\Delta \in \mathscr{L}$ and all (α, β),

$$R(\alpha,\beta,\Delta) = R(\beta,\alpha,\Delta). \tag{1.7}$$

Among the contributions to the two-armed bandit problem the work of W. Vogel deserves special mention. Considering the same set-up we do, he discussed a certain subclass of the class $\mathscr{L}$ in [4], and obtained asymptotic bounds for the minimax risk for $N \to \infty$ in [5]. Since we shall not be concerned with asymptotics in this paper, we state the following result without a formal proof: The lower bound for the asymptotic minimax risk for $N \to \infty$ obtained by Vogel in [5] may be raised by a factor $2^{\frac{1}{2}}$. This is proved by applying the same method that was used in [5] to the optimal symmetric strategy for $\alpha+\beta = 1$ that was discussed in [4]. Combining this lower bound with the upper bound given in [5] we find that the asymptotic minimax risk must be between $0.265\,N^{\frac{1}{2}}$ and $0.376\,N^{\frac{1}{2}}$.

In Section 2 we study the Bayes strategies in $\mathscr{D}$. By means of a certain recurrence relation we arrive at a complete characterization of these strategies, thus generalizing D. Feldman's well-known result in [3] for the case where the experimenter knows the values of α and β except for their order. In addition we obtain expressions for the Bayes risk of any prior distribution. Using these results we proceed to derive in Section 3 certain monotonicity properties of $\delta(m, k; n, l)$ for any admissible strategy Δ in $\mathscr{D}$. Though these relations may seem intuitively evident, one does well to remember that the two-armed bandit problem has been shown to defy intuition in many aspects (cf. [2]). In Section 4 we prove the existence of an admissible symmetric minimax-risk strategy having the monotonicity properties just mentioned. This fact to some degree facilitates the search for minimax-risk strategies, but even so, the algebra involved becomes progressively more complicated with increasing N and seems to remain prohibitive already for N as small as 5.

2. Bayes strategies. For $\Delta \in \mathscr{D}$ we consider the expected number of successes $E(S \mid \alpha, \beta, \Delta)$ as a function of the $\delta(m, k; n, l)$. Clearly, the dependence on each $\delta(m, k; n, l)$ is linear. We denote the coefficient of $\delta(m, k; n, l)$ in $E(S \mid \alpha, \beta, \Delta)$ (and hence also in $-R(\alpha, \beta, \Delta)$) by $p_\Delta(m, k; n, l) c_{\alpha,\beta,\Delta}(m, k; n, l)$. If all $\delta(m, k; n, l)$ are strictly between 0 and 1, then all $p_\Delta(m, k; n, l)$ are positive and as a result all $c_{\alpha,\beta,\Delta}(m, k; n, l)$ are uniquely determined. Otherwise the $c_{\alpha,\beta,\Delta}(m, k; n, l)$ are defined by continuity.

THEOREM 1. *For any strategy Δ in $\mathscr{D}$ the functions $c_{\alpha,\beta,\Delta}(m, k; n, l)$ satisfy the following relations*

$$c_{\alpha,\beta,\Delta}(m, k; n, l) = (\alpha-\beta)\alpha^k(1-\alpha)^{m-k}\beta^l(1-\beta)^{n-l} \tag{2.1}$$

if $m+n = N-1$,

$$\begin{aligned} c_{\alpha,\beta,\Delta}(m, k; n, l) = {} & \delta(m+1, k+1; n, l) c_{\alpha,\beta,\Delta}(m+1, k+1; n, l) \\ & + \delta(m+1, k; n, l) c_{\alpha,\beta,\Delta}(m+1, k; n, l) \\ & + [1-\delta(m, k; n+1, l+1)] c_{\alpha,\beta,\Delta}(m, k; n+1, l+1) \\ & + [1-\delta(m, k; n+1, l)] c_{\alpha,\beta,\Delta}(m, k; n+1, l) \end{aligned} \tag{2.2}$$

if $m+n \leqq N-2$.

PROOF. By continuity it is obviously sufficient to consider the case where all $\delta(m, k; n, l)$ as well as α and β are strictly between 0 and 1. This ensures that expression (1.3) is positive for all states $(m, k; n, l)$. Hence the conditional expectation $e_{\alpha,\beta,\Delta}(m, k; n, l)$ of the total number of successes S under α, β and Δ given that the state $(m, k; n, l)$ is reached, exists. It is clearly a linear function of $\delta(m, k; n, l)$ and may thus be written in the form

$$e_{\alpha,\beta,\Delta}(m, k; n, l) = a_{\alpha,\beta,\Delta}(m, k; n, l)\,\delta(m, k; n, l) + b_{\alpha,\beta,\Delta}(m, k; n, l). \tag{2.3}$$

It follows that

$$c_{\alpha,\beta,\Delta}(m, k; n, l) = a_{\alpha,\beta,\Delta}(m, k; n, l)\,\alpha^k(1-\alpha)^{m-k}\beta^l(1-\beta)^{n-l}. \tag{2.4}$$

Dropping the subscripts α, β and Δ, we obtain, from the definition of $e(m, k; n, l)$,

$$\begin{aligned} e(m, k; n, l) = {} & \delta(m, k; n, l)[\alpha e(m+1, k+1; n, l) + (1-\alpha)\, e(m+1, k; n, l)] \\ & + [1-\delta(m, k; n, l)][\beta e(m, k; n+1, l+1) \\ & \qquad + (1-\beta)\, e(m, k; n+1, l)], \end{aligned} \tag{2.5}$$

and consequently

$$a(m,k;n,l) = \alpha e(m+1,k+1;n,l)+(1-\alpha)e(m+1,k;n,l) \\ -\beta e(m,k;n+1,l+1)-(1-\beta)e(m,k;n+1,l), \tag{2.6}$$

$$b(m,k;n,l) = \beta e(m,k;n+1,l+1)+(1-\beta)e(m,k;n+1,l). \tag{2.7}$$

If $m+n = N-1$, then (2.6) becomes $a(m, k; n, l) = \alpha-\beta$, and hence (2.1) follows from (2.4). On the other hand, rewriting (2.6) by means of (2.3) leads to

$$\begin{aligned} a(m,k;n,l) = {} & \alpha\delta(m+1,k+1;n,l)a(m+1,k+1;n,l) \\ & +(1-\alpha)\delta(m+1,k;n,l)a(m+1,k;n,l) \\ & +\beta[1-\delta(m,k;n+1,l+1)]a(m,k;n+1,l+1) \\ & +(1-\beta)[1-\delta(m,k;n+1,l)]a(m,k;n+1,l) \\ & +[\alpha b(m+1,k+1;n,l)+(1-\alpha)b(m+1,k;n,l) \\ & -\beta b(m,k;n+1,l+1) \\ & -(1-\beta)b(m,k;n+1,l)-\beta a(m,k;n+1,l+1) \\ & -(1-\beta)a(m,k;n+1,l)], \end{aligned}$$

where for $m+n = N-2$ the last expression between square brackets vanishes as one easily verifies using (2.6) and (2.7). This result, combined with (2.4), gives (2.2).

Let μ be a prior distribution on the closed unit square. For a strategy $\Delta \in \mathscr{D}$,

$$\rho(\mu,\Delta) = \int R(\alpha,\beta,\Delta)\,d\mu(\alpha,\beta) \tag{2.8}$$

denotes the average risk of Δ against μ. If we define

$$\gamma_{\mu,\Delta}(m,k;n,l) = \int c_{\alpha,\beta,\Delta}(m,k;n,l)\,d\mu(\alpha,\beta), \tag{2.9}$$

then $-p_\Delta(m, k; n, l)\gamma_{\mu,\Delta}(m, k; n, l)$ is the coefficient of $\delta(m, k; n, l)$ in $\rho(\mu, \Delta)$. It follows that any strategy Δ that has $\delta(m, k; n, l) = 1$ whenever $\gamma_{\mu,\Delta}(m, k; n, l) > 0$ and $\delta(m, k; n, l) = 0$ whenever $\gamma_{\mu,\Delta}(m, k; n, l) < 0$, minimizes $\rho(\mu, \Delta)$ for fixed μ and is therefore a Bayes strategy against μ. This may be seen by successively finding the optimal $\delta(m, k; n, l)$ for $m+n = N-1, N-2, \cdots, 0$, and noting that for $m+n = \nu$ these optimal values do not depend on the values of $\delta(m, k; n, l)$ for $m+n < \nu$. Conversely, every Bayes strategy against μ has a version with $\delta(m, k; n, l) = 1$ (or 0) whenever $\gamma_{\mu,\Delta}(m, k; n, l) > 0$ (or < 0).

THEOREM 2. *Let μ be a prior distribution on the closed unit square and let $\gamma_\mu(m, k; n, l)$ be defined by*

$$\gamma_\mu(m,k;n,l) = \int (\alpha-\beta)\alpha^k(1-\alpha)^{m-k}\beta^l(1-\beta)^{n-l}\,d\mu(\alpha,\beta) \tag{2.10}$$

if $m+n = N-1$,

$$\gamma_\mu(m,k;n,l) = \gamma_\mu^+(m+1,k+1;n,l)+\gamma_\mu^+(m+1,k;n,l) \\ -\gamma_\mu^-(m,k;n+1,l+1)-\gamma_\mu^-(m,k;n+1,l) \tag{2.11}$$

for $m+n \leqq N-2$, *where* x^+ *and* x^- *denote* $\max(0, x)$ *and* $\max(0, -x)$ *respectively. Then* $\Delta \in \mathscr{D}$ *is a Bayes strategy against* μ *if and only if it has a version with* $\delta(m, k; n, l) = 1$ *whenever* $\gamma_\mu(m, k; n, l) > 0$ *and* $\delta(m, k; n, l) = 0$ *whenever* $\gamma_\mu(m, k; n, l) < 0$.

PROOF. According to the remarks preceding the theorem, Δ is Bayes against μ iff it has a version for which $\delta(m, k; n, l) = 1$ (or 0) if $\gamma_{\mu,\Delta}(m, k; n, l) > 0$ (or < 0). Integrating (2.1) and (2.2) with respect to μ and substituting the values of the $\delta(m, k; n, l)$ we find that for this version of Δ, $\gamma_{\mu,\Delta}(m, k; n, l)$ equals $\gamma_\mu(m, k; n, l)$ as defined by (2.10) and (2.11) for all states.

We note that D. Feldman's characterization of the Bayes strategies in $\mathscr{D}$ against a prior distribution μ, which puts mass ξ and $1-\xi$ at points (α_0, β_0) and (β_0, α_0) respectively (cf. [3]), may be formulated as follows: Δ in $\mathscr{D}$ is Bayes against μ iff it has a version for which $\delta(m, k; n, l) = 1$ whenever $\eta_\mu(m, k; n, l) > 0$ and $\delta(m, k; n, l) = 0$ whenever $\eta_\mu(m, k; n, l) < 0$ where

$$\eta_\mu(m,k;n,l) = \xi\alpha_0^{\,k}(1-\alpha_0)^{m-k}\beta_0^{\,l}(1-\beta_0)^{n-l} - (1-\xi)\alpha_0^{\,l}(1-\alpha_0)^{n-l}\beta_0^{\,k}(1-\beta_0)^{m-k}$$

for all states $(m, k; n, l)$. It follows that $\operatorname{sgn}\eta_\mu(m, k; n, l) = \operatorname{sgn}\gamma_\mu(m, k; n, l)$ for all states $(m, k; n, l)$ and all μ of the type considered by Feldman. This fact may also be verified by a direct, though somewhat tedious argument.

To conclude this section we consider the Bayes risk $\rho(\mu)$ of an arbitrary prior distribution μ. This is defined as the average risk $\rho(\mu, \Delta)$ of any Bayes strategy Δ against μ, or equivalently, $\rho(\mu) = \inf_{\Delta \in \mathscr{D}} \rho(\mu, \Delta)$.

THEOREM 3. *For any prior distribution* μ,

$$\begin{aligned}
\rho(\mu) &= N\int\frac{|\alpha-\beta|}{2}\,d\mu(\alpha,\beta) - \sum_{m=0}^{N-1}\sum_{n=0}^{N-m-1}\sum_{k=0}^{m}\sum_{l=0}^{n}\frac{\binom{m+n}{n}\binom{m}{k}\binom{n}{l}}{2^{m+n+1}}|\gamma_\mu(m,k;n,l)| \\
&= N\textstyle\int(\alpha-\beta)^+\,d\mu(\alpha,\beta) - \sum_{n=0}^{N-1}\sum_{l=0}^{n}\binom{n}{l}\gamma_\mu^{\,+}(0,0;n,l) \\
&= N\textstyle\int(\alpha-\beta)^-\,d\mu(\alpha,\beta) - \sum_{m=0}^{N-1}\sum_{k=0}^{m}\binom{m}{k}\gamma_\mu^{\,-}(m,k;0,0).
\end{aligned}$$

PROOF. Let $\Delta \in \mathscr{D}$ be Bayes against μ. Without loss of generality we may restrict attention to a version of Δ which has the property described in Theorem 2. For any such version and any state $(m, k; n, l)$ with $m+n \leqq N-1$ we have

$$\begin{aligned}
\gamma_{\mu,\Delta}(m,k;n,l) &= \gamma_\mu(m,k;n,l), \\
(\delta(m,k;n,l)-\tfrac{1}{2})\gamma_\mu(m,k;n,l) &= \tfrac{1}{2}|\gamma_\mu(m,k;n,l)|, \\
\delta(m,k;n,l)\gamma_\mu(m,k;n,l) &= \gamma_\mu^{\,+}(m,k;n,l), \\
-(1-\delta(m,k;n,l))\gamma_\mu(m,k;n,l) &= \gamma_\mu^{\,-}(m,k;n,l).
\end{aligned}$$

Consequently for any state $(m, k; n, l)$ with $m+n \leqq N-1$ we obtain the following equalities, using (2.5) and the fact that $\gamma_{\mu,\Delta}(m, k; n, l)$ and hence $\gamma_\mu(m, k; n, l)$ equals the coefficient of $\delta(m, k; n, l)$ in the first member:

$$\begin{aligned}
&\int \alpha^k(1-\alpha)^{m-k}\beta^l(1-\beta)^{n-l}\, e_{\alpha,\beta,\Delta}(m, k; n, l)\, d\mu(\alpha, \beta)\\
&\quad = \tfrac{1}{2}|\gamma_\mu(m, k; n, l)|\\
&\qquad + \tfrac{1}{2}\int \alpha^{k+1}(1-\alpha)^{m-k}\beta^l(1-\beta)^{n-l}\, e_{\alpha,\beta,\Delta}(m+1, k+1; n, l)\, d\mu(\alpha, \beta)\\
&\qquad + \tfrac{1}{2}\int \alpha^k(1-\alpha)^{m-k+1}\beta^l(1-\beta)^{n-l}\, e_{\alpha,\beta,\Delta}(m+1, k; n, l)\, d\mu(\alpha, \beta)\\
&\qquad + \tfrac{1}{2}\int \alpha^k(1-\alpha)^{m-k}\beta^{l+1}(1-\beta)^{n-l}\, e_{\alpha,\beta,\Delta}(m, k; n+1, l+1)\, d\mu(\alpha, \beta)\\
&\qquad + \tfrac{1}{2}\int \alpha^k(1-\alpha)^{m-k}\beta^l(1-\beta)^{n-l+1}\, e_{\alpha,\beta,\Delta}(m, k; n+1, l)\, d\mu(\alpha, \beta)\\
&\quad = \gamma_\mu^{+}(m, k; n, l) \qquad\qquad (2.12)\\
&\qquad + \int \alpha^k(1-\alpha)^{m-k}\beta^{l+1}(1-\beta)^{n-l}\, e_{\alpha,\beta,\Delta}(m, k; n+1, l+1)\, d\mu(\alpha, \beta)\\
&\qquad + \int \alpha^k(1-\alpha)^{m-k}\beta^l(1-\beta)^{n-l+1}\, e_{\alpha,\beta,\Delta}(m, k; n+1, l)\, d\mu(\alpha, \beta)\\
&\quad = \gamma_\mu^{-}(m, k; n, l)\\
&\qquad + \int \alpha^{k+1}(1-\alpha)^{m-k}\beta^l(1-\beta)^{n-l}\, e_{\alpha,\beta,\Delta}(m+1, k+1; n, l)\, d\mu(\alpha, \beta)\\
&\qquad + \int \alpha^k(1-\alpha)^{m-k+1}\beta^l(1-\beta)^{n-l}\, e_{\alpha,\beta,\Delta}(m+1, k; n, l)\, d\mu(\alpha, \beta).
\end{aligned}$$

Observing that by definition $E(S \mid \alpha, \beta, \Delta) = e_{\alpha,\beta,\Delta}(0, 0; 0, 0)$ and $e_{\alpha,\beta,\Delta}(m, k; n, l) = k+l$ for any state $(m, k; n, l)$ with $m+n = N$, we arrive at the three desired expressions by repeated application of the corresponding versions of (2.12).

3. Admissible strategies. For the type of problem considered in this paper every admissible strategy is also a Bayes strategy. In the sequel we shall, however, need a slightly stronger result. We shall say that a prior distribution is nonmarginal if, for some $\varepsilon > 0$, it assigns probability 1 to the set

$$Q_\varepsilon = \{(\alpha, \beta) \mid |\alpha-\beta|\,\alpha(1-\alpha)\,\beta(1-\beta) \geqq \varepsilon, 0 < \alpha < 1, 0 < \beta < 1\}. \tag{3.1}$$

THEOREM 4. *Every admissible strategy $\Delta \in \mathscr{D}$ is Bayes against a nonmarginal prior distribution.*

PROOF. Let Δ be any strategy which is not Bayes against any nonmarginal prior. It is sufficient to show that Δ is not admissible.

For any sufficiently small $\varepsilon_i > 0$, consider the restricted problem where the parameter space is reduced to the set $A_i = Q_{\varepsilon_i}$ as defined by (3.1). Since A_i is compact, the assertion that every admissible strategy is Bayes remains true for the restricted problem. By our assumption Δ is not Bayes, and therefore not admissible in the new problem. It follows that there exists a strategy Δ_i that is Bayes against a prior distribution μ_i on A_i and for which $R(\alpha, \beta, \Delta_i) \leqq R(\alpha, \beta, \Delta)$ for all $(\alpha, \beta) \in A_i$. By a standard procedure we may select a sequence $\varepsilon_i \searrow 0$ and corresponding μ_i and

Δ_i such that the strategies Δ_i converge to a strategy Δ_0 in the sense that $\delta_i(m, k; n, l)$ converges to $\delta_0(m, k; n, l)$ for every state $(m, k; n, l)$. Obviously

$$R(\alpha, \beta, \Delta_0) \leqq R(\alpha, \beta, \Delta) \qquad \text{for all} \quad \alpha, \beta \in [0, 1]$$

since the inequality must hold on every A_i and both functions are continuous.

Since Δ_i converges to Δ_0 there exists a positive integer j for which Δ_j has the following properties:

(a) For all states with $\delta_0(m, k; n, l) = 0$, $\delta_j(m, k; n, l) \neq 1$;
(b) For all states with $\delta_0(m, k; n, l) = 1$, $\delta_j(m, k; n, l) \neq 0$;
(c) For all states with $0 < \delta_0(m, k; n, l) < 1$, $0 < \delta_j(m, k; n, l) < 1$.

This implies that $\delta_0(m, k; n, l) = \delta_j(m, k; n, l)$ for every state with $\delta_j(m, k; n, l) = 0$ or 1. Recalling that Δ_j is Bayes against μ_j and noting that this property can not be destroyed by changing only those $\delta_j(m, k; n, l)$ that are strictly between 0 and 1, we find that Δ_0 is Bayes against the prior distribution μ_j on A_j. As Δ is not Bayes against μ_j by our assumption, the inequality $R(\alpha, \beta, \Delta_0) \leqq R(\alpha, \beta, \Delta)$ on the closed unit square must be strict for at least one point (α, β) and the inadmissibility of Δ follows.

We are now in a position to prove a theorem that provides some insight in the structure of admissible strategies.

THEOREM 5. *If μ is a nonmarginal prior distribution and $m+n \leqq N-2$, then*

$$\gamma_\mu(m, k; n+1, l+1) < \gamma_\mu(m+1, k+1; n, l) \tag{3.2}$$

$$\gamma_\mu(m+1, k; n, l) < \gamma_\mu(m, k; n+1, l) \tag{3.3}$$

PROOF. For $m+n = N-2$, (2.10) yields

$$\gamma_\mu(m+1, k+1; n, l) - \gamma_\mu(m, k; n+1, l+1) = \int (\alpha-\beta)^2 \alpha^k (1-\alpha)^{m-k} \beta^l (1-\beta)^{n-l} \, d\mu(\alpha, \beta),$$

which is strictly positive since μ is nonmarginal. In the same way one shows that (3.3) is satisfied for $m+n = N-2$.

Next we suppose that the theorem is valid for $m+n = \nu$, where $0 < \nu \leqq N-2$, and we assume $m+n = \nu-1$. By (2.11) we have then

$$\begin{aligned}
\gamma_\mu(m+1, k+1; n, l) &- \gamma_\mu(m, k; n+1, l+1) \\
&= [\gamma_\mu^{+}(m+2, k+2; n, l) - \gamma_\mu^{+}(m+1, k+1; n+1, l+1)] \\
&\quad + [\gamma_\mu^{+}(m+2, k+1; n, l) - \gamma_\mu^{+}(m+1, k; n+1, l+1)] \\
&\quad + [\gamma_\mu^{-}(m, k; n+2, l+2) - \gamma_\mu^{-}(m+1, k+1; n+1, l+1)] \\
&\quad + [\gamma_\mu^{-}(m, k; n+2, l+1) - \gamma_\mu^{-}(m+1, k+1, n+1, l)] \geqq 0
\end{aligned}$$

since by hypothesis each of these four expressions is nonnegative. Equality can occur only if all four expressions vanish. However, the first and the third one can vanish only if $\gamma_\mu(m+1, k+1; n+1, l+1) < 0$ and $\geqq 0$ respectively, and hence inequality (3.2) is strict.

Similarly (3.3) follows from

$$\begin{aligned}\gamma_\mu(m, k; n+1, l) - \gamma_\mu(m+1, k; n, l) \\ &= [\gamma_\mu^{+}(m+1, k+1; n+1, l) - \gamma_\mu^{+}(m+2, k+1; n, l)] \\ &\quad + [\gamma_\mu^{+}(m+1, k; n+1, l) - \gamma_\mu^{+}(m+2, k; n, l)] \\ &\quad + [\gamma_\mu^{-}(m+1, k; n+1, l+1) - \gamma_\mu^{-}(m, k; n+2, l+1)] \\ &\quad + [\gamma_\mu^{-}(m+1, k; n+1, l) - \gamma_\mu^{-}(m, k; n+2, l)] \geqq 0\end{aligned}$$

and the fact that the first expression in square brackets can vanish only if $\gamma_\mu(m+2, k+1; n, l) < 0$ and the third one only if $\gamma_\mu(m+1, k; n+1, l+1) \geqq 0$, which would imply $\gamma_\mu(m+2, k+1; n, l) > 0$.

COROLLARY 1. *Every admissible strategy $\Delta \in \mathscr{D}$ has a version for which*

$$\delta(m, k; n+1, l+1) \leqq \delta(m+1, k+1; n, l) \tag{3.4}$$

$$\delta(m+1, k; n, l) \leqq \delta(m, k; n+1, l) \tag{3.5}$$

for all $m+n \leqq N-2$, where in each of these inequalities at least one member equals 0 or 1.

PROOF. By Theorem 4, Δ is Bayes against a nonmarginal prior μ, and as a result the theorem is proved by applying Theorem 5 and Theorem 2.

COROLLARY 2. *Every admissible strategy $\Delta \in \mathscr{D}$ has a version for which*

$$\delta(m, k; n, l)[1-\delta(m+1, k+1; n, l)][1-\delta(m+1, k; n, l)] = 0 \tag{3.6}$$

$$[1-\delta(m, k; n, l)]\,\delta(m, k; n+1, l+1)\,\delta(m, k; n+1, l) = 0 \tag{3.7}$$

for all $m+n \leqq N-2$.

PROOF. As before, we let μ denote the nonmarginal prior of Theorem 4 and consider the version of Δ having $\delta(m, k; n, l) = 1$ (or 0) whenever $\gamma_\mu(m, k; n, l) > 0$ (or < 0). If (3.6) were false for this version, then $\gamma_\mu(m, k; n, l) \geqq 0$, $\gamma_\mu(m+1, k+1; n, l) \leqq 0$ and $\gamma_\mu(m+1, k; n, l) \leqq 0$. The second of these inequalities implies $\gamma_\mu(m, k; n+1, l+1) < 0$ by Theorem 5, and hence (2.11) shows that $\gamma_\mu(m, k; n, l) < 0$, which contradicts the first inequality.

Similarly, if (3.7) were false, then $\gamma_\mu(m, k; n, l) \leqq 0$, $\gamma_\mu(m, k; n+1, l+1) \geqq 0$ and $\gamma_\mu(m, k; n+1, l) \geqq 0$. The second inequality implies $\gamma_\mu(m+1, k+1; n, l) > 0$ by Theorem 5, and hence $\gamma_\mu(m, k; n, l) > 0$ by (2.11), which contradicts the first inequality.

Intuitively one might expect some further monotonicity relations, like e.g. (i): $\delta(m, k; n, l) \leqq \delta(m+1, k+1; n, l)$ and (ii): $\delta(m, k; n, l) \leqq \delta(m, k+1; n, l)$, for any reasonable strategy in $\mathscr{D}$. However, (i) is nothing but another version of Bradt, Johnson and Karlin's principle of staying on a winner (cf. [2]), which they showed not to be generally true for all Bayes strategies in $\mathscr{D}$. In fact, (i) and (ii) do not even hold for all admissible strategies in $\mathscr{D}$ as one can see from the example given in [2]: The Bayes strategies in $\mathscr{D}$ for the case $N = 2$ against the prior distribution μ, which puts mass .8 in (.1, 0) and mass .2 in (.9, 1), are precisely those strategies in $\mathscr{D}$ for which $\delta(0, 0; 0, 0) = 1$, $\delta(1, 1; 0, 0) = 0$, and $\delta(1, 0; 0, 0) = 1$. Thus there is an essentially unique and hence admissible Bayes strategy against μ, which violates (i) and (ii).

For admissible strategies, which are also symmetric, Corollary 1 takes the following more explicit form.

COROLLARY 3. *Every admissible strategy $\Delta \in \mathscr{L}$ has a version for which*

$$\delta(m,k;n,l) = 1, \qquad \delta(n,l;m,k) = 0 \tag{3.8}$$

whenever $m+n \leqq N-1, k \geqq 1, m-k \leqq n-l$ and $(m,k;n,l) \neq (n,l;m,k)$.

PROOF. For the version of Δ that satisfies Corollary 1 we find by repeated application of (3.4) and (3.5) $\delta(m, k; n, l) \geqq \delta(m-k+l, l; n+k-l, k) \geqq \delta(n, l; m, k)$ where at least one of the extreme members must be 0 or 1. Since their sum equals 1 if $p_\Delta(m, k; n, l) \neq 0$, (3.8) will hold in this case. If $p_\Delta(m, k; n, l) = 0$, then by (1.6) we also have $p_\Delta(n, l; m, k) = 0$ and choosing $\delta(m, k; n, l) = 1$ and $\delta(n, l; m, k) = 0$ merely leads to another version of Δ.

We conclude this section by remarking that Corollaries 1, 2 and 3 obviously continue to hold if, instead of admissibility, we require that Δ be Bayes against a nonmarginal prior.

4. Symmetric minimax-risk strategies.

THEOREM 6. *There is a minimax-risk strategy which is admissible and belongs to $\mathscr{L}$.*

PROOF. The class $\mathscr{D}$, with the topology induced by the notion of convergence introduced in the proof of Theorem 4, is compact. The existence of a minimax-risk strategy in $\mathscr{D}$ is a well-known consequence of this. Moreover, the class $\mathscr{D}^*$ of all minimax-risk strategies in $\mathscr{D}$ is easily seen to be closed. Thus, if ν denotes Lebesgue measure on the unit square, there is a strategy $\Delta_1 \in \mathscr{D}^*$ such that $\rho(\nu, \Delta_1) = \min_{\Delta \in \mathscr{D}^*} \rho(\nu, \Delta)$. This follows from the continuity of $\rho(\nu, \cdot)$. Let $\Delta_2 \in \mathscr{D}$ be defined by $\delta_2(m, k; n, l) = 1-\delta_1(n, l; m, k)$ for all states $(m, k; n, l)$. Then $p_{\Delta_2}(m, k; n, l) = p_{\Delta_1}(n, l; m, k)$ for all states, and hence $R(\alpha, \beta, \Delta_2) = R(\beta, \alpha, \Delta_1)$ for all (α, β), so that $\Delta_2 \in \mathscr{D}^*$. By convexity we now may construct a strategy $\Delta \in \mathscr{D}$ satisfying (1.4) with $\lambda = \frac{1}{2}$. It follows that $R(\alpha, \beta, \Delta) = \frac{1}{2}R(\alpha, \beta, \Delta_1)+\frac{1}{2}R(\alpha, \beta, \Delta_2)$ for all (α, β), and hence $\Delta \in \mathscr{D}^*$. Finally we define $\Delta^* \in \mathscr{L}$ by

$$\delta^*(m,k;n,l) = \tfrac{1}{2}\delta(m,k;n,l)+\tfrac{1}{2}[1-\delta(n,l;m,k)]$$

for all states. The construction of Δ implies that $p_{\Delta^*}(m, k; n, l) = p_\Delta(m, k; n, l)$ for all states, and hence $\Delta^* \in \mathscr{D}^* \cap \mathscr{L}$.

In order to show that Δ^* is also admissible, we first remark that any strategy outside $\mathscr{D}^*$ has at some point (α, β) strictly larger risk than Δ^*, because Δ^* has minimax-risk. On the other hand, going through the steps leading to the construction of Δ^* once more, one easily verifies that $\rho(v, \Delta_1) = \rho(v, \Delta_2) = \rho(v, \Delta) = \rho(v, \Delta^*)$, so that $\rho(v, \Delta^*) \leqq \rho(v, \Delta')$ for any $\Delta' \in \mathscr{D}^*$. But because of the continuity of $R(\cdot, \cdot, \Delta)$, this implies that also within $\mathscr{D}^*$ there is no strategy improving on Δ^*, and thus the proof is complete.

The above proof really consists of two separate arguments mixed together. The first one is quite standard (cf. e.g. Theorem 8.6.4. in [1] and shows the existence of a symmetric minimax-risk strategy. The second argument, yielding admissibility, exploits an idea of Wald ([6] page 102). By the same argument, replacing $\mathscr{D}^*$ by the class of all Bayes strategies against any given prior distribution μ, one can prove the existence of an admissible Bayes strategy against μ.

Theorem 6 together with Corollaries 1, 2 and 3 yields

COROLLARY 4. *There is an admissible symmetric minimax-risk strategy which obeys* (3.4) *through* (3.8).

For $N = 1$ or 2, (1.5) and (3.8) uniquely determine a symmetric strategy. It follows from Corollary 4 and Corollary 3 that this strategy has minimax risk and is in fact the only admissible strategy in $\mathscr{L}$. For $N \geqq 3$ the situation rapidly becomes more complicated. In order to find a symmetric minimax-risk strategy Δ_0 satisfying (3.4) through (3.8) one first has to find a general expression for the risk function $R(\alpha, \beta, \Delta)$ of an arbitrary symmetric strategy Δ satisfying (3.8). Then, with the aid of (3.4) through (3.7), one has to solve the remaining $\delta(m, k; n, l)$ directly using the minimax property.

To accomplish the first step of computing $R(\alpha, \beta, \Delta)$ for an arbitrary symmetric strategy, one may proceed recursively. This is especially useful if one wants to find $R(\alpha, \beta, \Delta)$ for a number of values of N. If $X_v = 1 - Y_v = 1$ or 0 according to whether E_1 or E_2 is carried out on the vth trial ($v = 1, 2, \cdots, N$), then $R(\alpha, \beta, \Delta)$, being equal to $|\alpha - \beta|$ multiplied by the expected number of times the experimenter uses the less favorable experiment, is given by

$$R(\alpha, \beta, \Delta) = \tfrac{1}{2}N\,|\alpha - \beta| - \tfrac{1}{2}(\alpha - \beta)\textstyle\sum_{v=1}^{N} E(X_v - Y_v \,|\, \alpha, \beta, \Delta). \tag{4.1}$$

Remembering the definition of $\pi_{\alpha,\beta,\Delta}(m, k; n, l)$, we have

$$E(X_v - Y_v \,|\, \alpha, \beta, \Delta) = \textstyle\sum \pi_{\alpha,\beta,\Delta}(m, k; n, l)[2\delta(m, k; n, l) - 1], \tag{4.2}$$

where the summation is extended over all states $(m, k; n, l)$ with $m + n = v - 1$, and where the $\pi_{\alpha,\beta,\Delta}(m, k; n, l)$ can be computed recursively by means of

$$\begin{aligned} \pi_{\alpha,\beta,\Delta}(m, k; n, l) = {} & \alpha\delta(m-1, k-1; n, l)\pi_{\alpha,\beta,\Delta}(m-1, k-1; n, l) \\ & + (1-\alpha)\,\delta(m-1, k; n, l)\,\pi_{\alpha,\beta,\Delta}(m-1, k; n, l) \\ & + \beta[1 - \delta(m, k; n-1, l-1)]\,\pi_{\alpha,\beta,\Delta}(m, k; n-1, l-1) \\ & + (1-\beta)[1 - \delta(m, k; n-1, l)]\,\pi_{\alpha,\beta,\Delta}(m, k; n-1, l) \end{aligned} \tag{4.3}$$

starting from

$$\begin{aligned} \pi_{\alpha,\beta,\Delta}(0, k; 0, l) &= 1 \qquad \text{if } k = l = 0; \\ &= 0 \qquad \text{otherwise.} \end{aligned} \tag{4.4}$$

The work involved may be reduced somewhat by means of the relation

$$\pi_{\alpha,\beta,\Delta}(m,k;n,l) = \pi_{\alpha,\beta,\Delta}(n,l;m,k), \tag{4.5}$$

which is a consequence of (1.3) and (1.6).

For $N=3$, only $\delta(2, 1; 0, 0)$ remains undetermined by the requirement that Δ be symmetric and must satisfy (3.8), and one finds

$$R(\alpha,\beta,\Delta) = \tfrac{3}{2}|\alpha-\beta| - \tfrac{1}{4}(\alpha-\beta)^2\{1+\delta(2,1;0,0)+[1-\delta(2,1;0,0)](\alpha+\beta)\}.$$

After a little algebra one sees that Δ_0 must have $\delta(2, 1; 0, 0) = 1$ and that $R(\alpha, \beta, \Delta_0)$ attains its maximum $M(\Delta_0) = \frac{9}{16}$ when $|\alpha-\beta| = \frac{3}{4}$.

For $N=4$ only $\delta(2, 1; 0, 0)$, $\delta(3, 1; 0, 0)$ and $\delta(3, 2; 0, 0)$ are to be determined and

$$\begin{aligned} R(\alpha,\beta,\Delta) = 2|\alpha-\beta| - \tfrac{1}{4}(\alpha-\beta)^2\{&(\alpha^2+\beta^2+3\alpha\beta-\alpha-\beta+3)-\delta(2,1;0,0)\alpha\beta \\ &-\delta(3,2;0,0)[1+\delta(2,1;0,0)](\alpha^2+\beta^2+\alpha\beta-\alpha-\beta) \\ &+\delta(3,1;0,0)\,\delta(2,1;0,0)(\alpha^2+\beta^2+\alpha\beta-2\alpha-2\beta+1)\}. \end{aligned}$$

Using (3.6), one finds after lengthy calculations that Δ_0 must have $\delta(2, 1; 0, 0) = \frac{4}{5}$, $\delta(3, 1; 0, 0) = \frac{1}{2}$ and $\delta(3, 2; 0, 0) = 1$, so that the risk function of Δ_0 is given by

$$R(\alpha,\beta,\Delta_0) = 2|\alpha-\beta| - \tfrac{17}{10}(\alpha-\beta)^2 + \tfrac{1}{5}(\alpha-\beta)^4$$

and attains its maximum $M(\Delta_0) = .617$ when $|\alpha-\beta| = .654$. For larger values of N the number of $\delta(m, k; n, 1)$ that have to be determined increases rapidly, and consequently the algebra involved becomes distressingly complicated.

REFERENCES

[1] BLACKWELL, D. and GIRSHICK, M. A. (1954). *Theory of Games and Statistical Decisions.* Wiley, New York.

[2] BRADT, R. N., JOHNSON, S. M. and KARLIN, S. (1956). On sequential designs for maximizing the sum of n observations. *Ann. Math. Statist.* **27** 1060–1074.

[3] FELDMAN, D. (1962). Contributions to the "two-armed bandit" problem. *Ann. Math. Statist.* **33** 847–856.

[4] VOGEL, W. (1960a). Ein Irrfahrten-Problem und seine Anwendung auf die Theorie der sequentiellen Versuchs-Pläne, *Arch. Math.* **11** 310–320.

[5] VOGEL, W. (1960b). An asymptotic minimax theorem for the two-armed bandit problem. *Ann. Math. Statist.* **31** 444–451.

[6] WALD, A. (1950). *Statistical Decision Functions.* Wiley, New York.